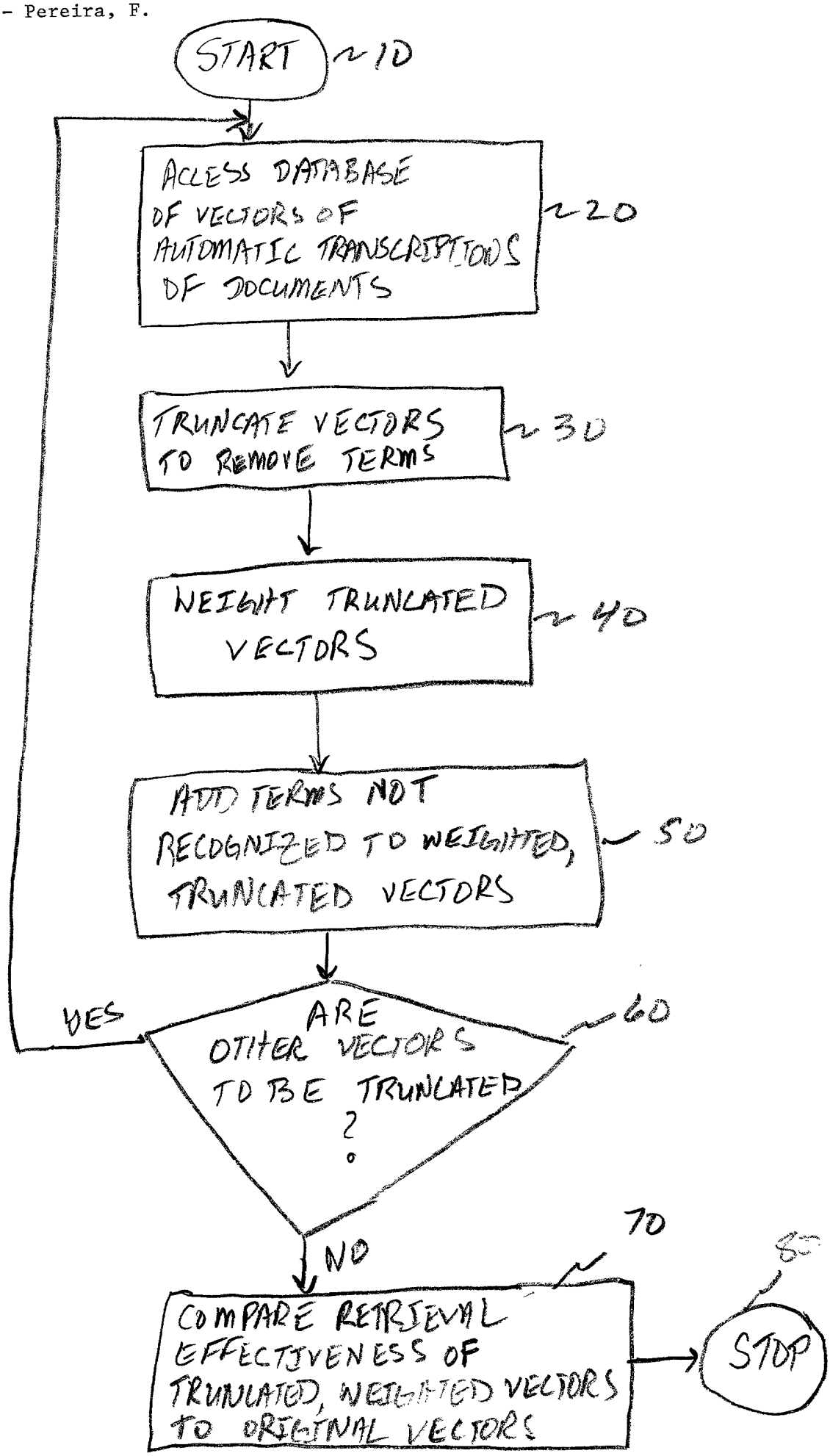

START
10
ACCESS DATABASE OF VECTORS OF AUTOMATIC TRANSCRIPTIONS OF DOCUMENTS
20
TRUNCATE VECTORS TO REMOVE TERMS
30
WEIGHT TRUNCATED VECTORS
40
ADD TERMS NOT RECOGNIZED TO WEIGHTED, TRUNCATED VECTORS
50
ARE OTHER VECTORS TO BE TRUNCATED ?
60
YES
NO
COMPARE RETRIEVAL EFFECTIVENESS OF TRUNCATED, WEIGHTED VECTORS TO ORIGINAL VECTORS
70
STOP
80

Fig. 1

Code	Provided By	WER
Human	NIST	0%
CUHTK-S1	Cambridge University	24.8%
Dragon98-S1	Dragon Systems	29.8%
ATT-S1	AT&T Labs	31.0%
NIST-B1	Carnegie Melon (CMU)	34.1%
SHEF-S1	Sheffield University	36.8%
NIST-B2	Carnegie Mellon (CMU)	46.9%
DERASRU-S2	DERA	61.5%
DERASRU-S1	DERA	66.2%

Table 1: Transcriptions used in this study.

Fig. 2

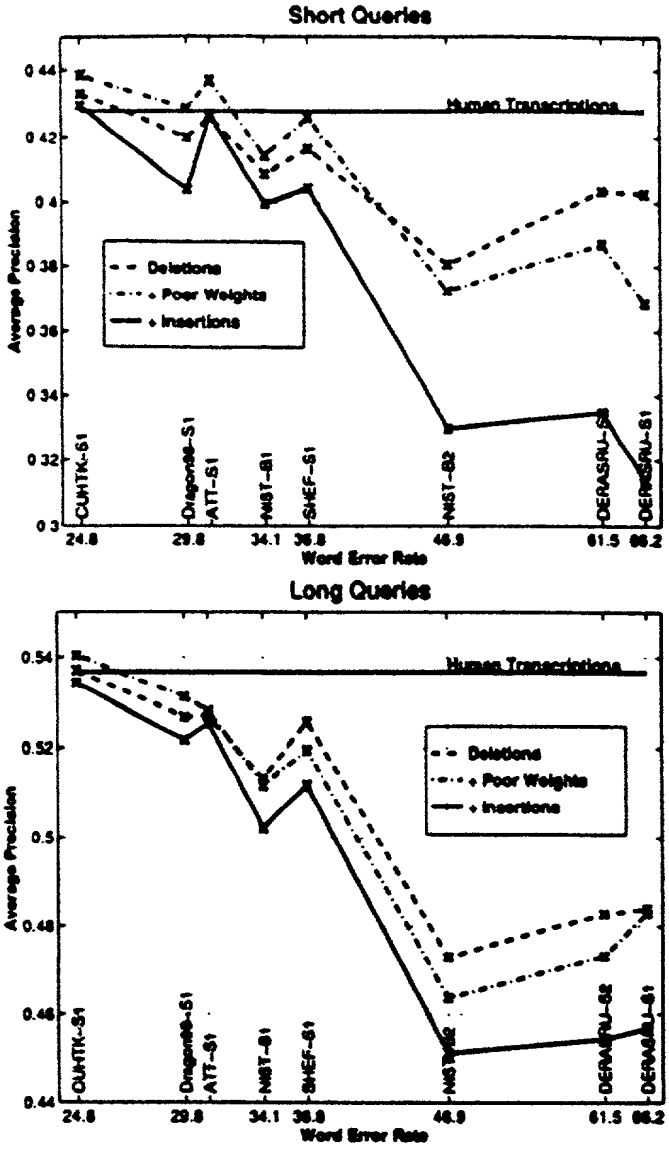


Figure 1: Losses due to different recognizer mistakes.

Fig. 3

(3)

Degree of Expansion	α in Rocchio			
	0.5	1.0	1.5	2.0
0%	0.4739	0.4656	0.4591	0.4556
10%	0.4898	0.4834	0.4756	0.4708
20%	0.4938	0.4866	0.4813	0.4763
50%	0.4990	0.4977	0.4898	0.4800
100%	0.4992	**0.4995**	0.4952	0.4892
200%	0.4943	0.4974	0.4923	0.4858

Fig. 4

Table 2: Parameter exploration for NIST-B1 (short queries).

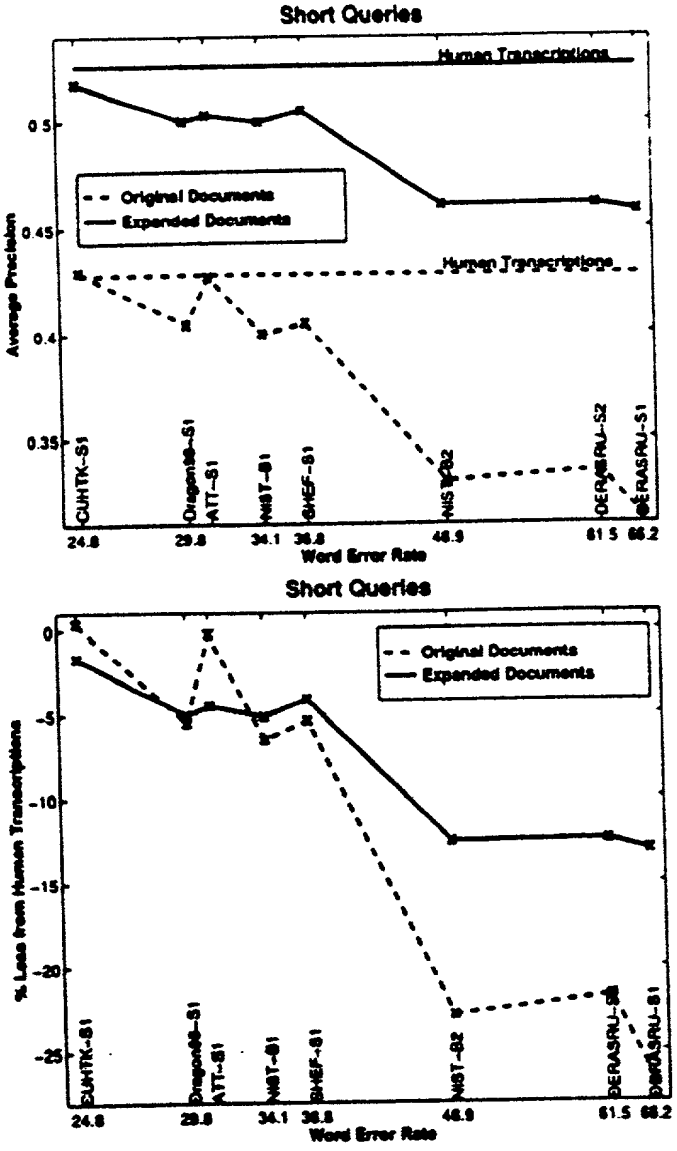
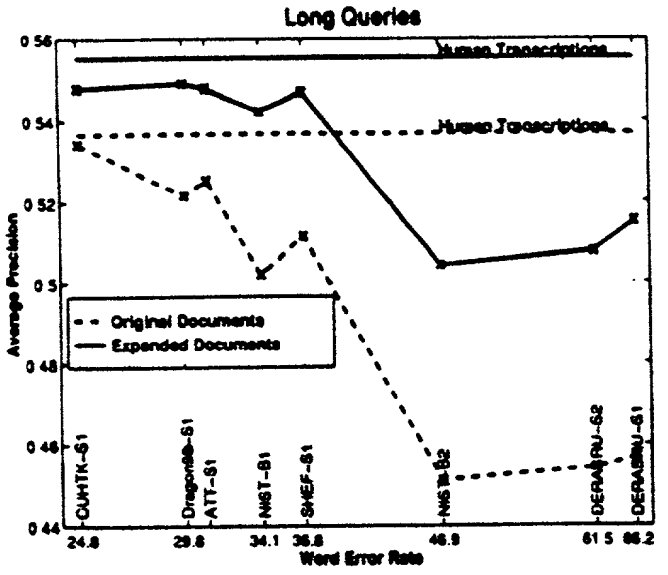


Figure 2: Document expansion for short queries.

Fig. 5


Long Queries
Human Transcriptions
Human Transcriptions
Original Documents
Expended Documents
Average Precision
0.56
0.54
0.52
0.5
0.48
0.46
0.44
CUHTK-S1
Dragon98-S1
ATT-S1
NIST-S1
SHEF-S1
DERASRU-S2
DERASRU-S1
24.8
29.8
34.1
36.8
46.9
61.5
66.2
Word Error Rate

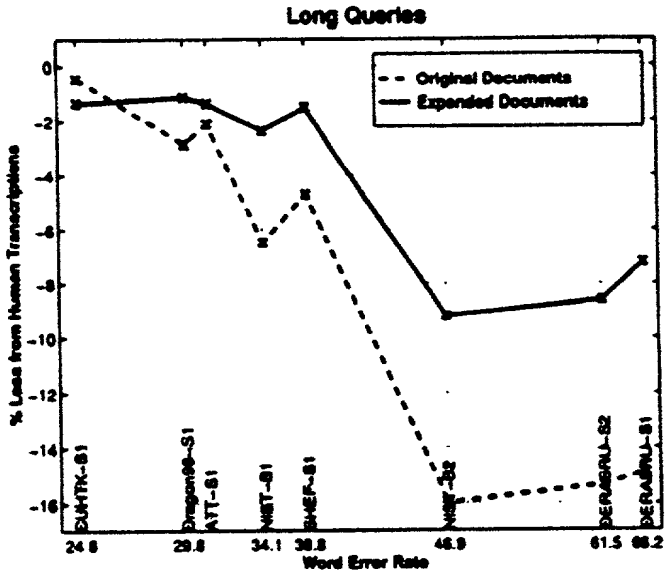
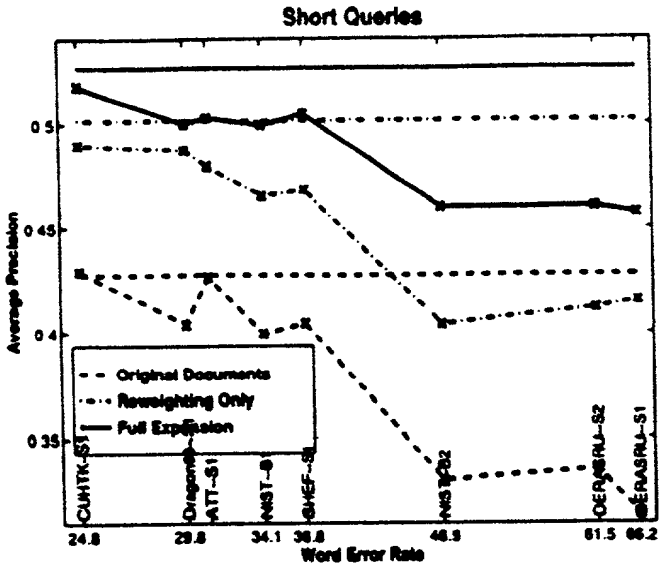

Long Queries
Original Documents
Expended Documents
% Loss from Human Transcriptions
0
-2
-4
-6
-8
-10
-12
-14
-16
CUHTK-S1
Dragon98-S1
ATT-S1
NIST-S1
SHEF-S1
DERASRU-S2
DERASRU-S1
24.8
29.8
34.1
36.8
46.9
61.5
66.2
Word Error Rate

Fig. 6


Short Queries
Average Precision
Word Error Rate
Original Documents
Reweighting Only
Full Expansion
0.5
0.45
0.4
0.35
24.8
29.8
34.1
36.8
46.9
61.5
66.2

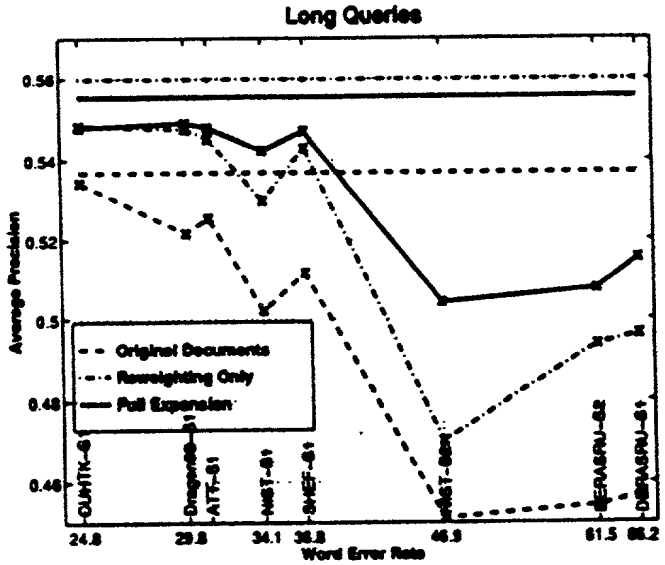

Long Queries
Average Precision
Word Error Rate
Original Documents
Reweighting Only
Full Expansion
0.56
0.54
0.52
0.5
0.48
0.46
24.8
29.8
34.1
36.8
46.9
61.5
66.2

Fig. 7